EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	4	"20040088308" or "20020087310"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/24 18:06
S2	2087	(document with search with query) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 19:38
S3	264	(document with search with query) and ((stop$4 or end$3) adj word) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 19:40
S4	48	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 19:41
S5	34	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and pars$5 and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 19:42
S6	44	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 19:42
S7	42	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and compar$5 and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 19:43
S8	0	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and compar$5 and @ad < "20040331" and (re$1writ$5 with query)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 19:43

EAST Search History

S9	14	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and compar$5 and @ad < "20040331" and (re$1writ$5 and query)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 19:44
S10	0	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and compar$5 and @ad < "20040331" and (re$1writ$5 same query)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 19:44
S11	6	"20040088308" or "20030088562" or "20030069877"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/15 20:21
S12	496	(modif$4 with query) and ((remov$4 or exclu$5) with (term or word))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/16 13:29
S13	34	(modif$4 with query) with ((remov$4 or exclu$5 or eliminat$4) with (term or word))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/16 13:29
S14	31	(modif$4 with query) with ((remov$4 or exclu$5 or eliminat$4) with (term or word)) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/17 12:16
S15	13	(modif$4 with query) and ((remov$4 or exclu$5 or eliminat$4) with (stop$3word or stop$3term)) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/16 15:34
S16	4	"20030233618" or "20030004914"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/16 15:35
S17	31	(modif$4 with query) with ((remov$4 or exclu$5 or eliminat$4) with (term or word)) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/17 12:16

S18	15	707/3.ccls. and S17	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/17 12:16
S19	1	707/102.ccls. and S17	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/17 12:16
S20	2	707/100.ccls. and S17	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/17 12:17
S21	0	707/101.ccls. and S17	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/17 12:17
S22	2	715/513.ccls. and S17	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/09/17 12:17
S23	2	query with (stop$1word or stop$1term or ((irrelavant or non$essential) adj (word or term))) with (identif$4 or detect)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/24 18:08
S24	2	query with (stop$1word or stop$1term or ((irrelavant or non$essential or non$1material) adj (phrase or word or term))) with (identif$4 or detect or filter)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 11:41
S25	2	query with (stop$1word or stop$1term or ((irrelavant or non$essential or non$1material or non$1interesting) adj (phrase or word or term)) or not$1word) with (identif$4 or detect or filter)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 11:42
S26	4	"20040088308" or "20020087310"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46

EAST Search History

S27	2179	(document with search with query) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S28	279	(document with search with query) and ((stop$4 or end$3) adj word) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S29	53	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S30	38	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and pars$5 and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S31	49	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S32	47	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and compar$5 and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S33	0	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and compar$5 and @ad < "20040331" and (re$1writ$5 with query)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S34	0	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and compar$5 and @ad < "20040331" and (re$1writ$5 same query)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46

S35	17	(document with search with query) and ((detect$4 or identif$5 or classif$7) with ((stop$4 or end$3) adj word)) and (analy$6 or pars$5) and compar$5 and @ad < "20040331" and (re$1writ$5 and query)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S36	6	"20040088308" or "20030088562" or "20030069877"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S37	550	(modif$4 with query) and ((remov$4 or exclu$5) with (term or word))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S38	44	(modif$4 with query) with ((remov$4 or exclu$5 or eliminat$4) with (term or word))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S39	32	(modif$4 with query) with ((remov$4 or exclu$5 or eliminat$4) with (term or word)) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S40	13	(modif$4 with query) and ((remov$4 or exclu$5 or eliminat$4) with (stop$3word or stop$3term)) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB ·	OR	ON	2007/02/25 13:46
S41	4	"20030233618" or "20030004914"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S42	32	(modif$4 with query) with ((remov$4 or exclu$5 or eliminat$4) with (term or word)) and @ad < "20040331"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S43	15	707/3.ccls. and S42	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46

S44	1	707/102.ccls. and S42	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S45	2	707/100.ccls. and S42	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S46	0	707/101.ccls. and S42	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S47	2	715/513.ccls. and S42	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S48	2	query with (stop$1word or stop$1term or ((irrelavant or non$essential) adj (word or term))) with (identif$4 or detect)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S49	2	query with (stop$1word or stop$1term or ((irrelavant or non$essential or non$1material) adj (phrase or word or term))) with (identif$4 or detect or filter)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46
S50	2	query with (stop$1word or stop$1term or ((irrelavant or non$essential or non$1material or non$1interesting) adj (phrase or word or term)) or not$1word) with (identif$4 or detect or filter)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2007/02/25 13:46

P☀RTAL

USPTO

Search: ◉ The ACM Digital Library   ○ The Guide

+query +stopwords +results +compare +stop +phrases +iden   **SEARCH**

THE ACM DIGITAL LIBRARY

Feedback Report a problem Satisfaction survey

Terms used
query stopwords results compare stop phrases identif

Found **1** of **197,895**

| Sort results by | relevance ▼ |
| Display results | expanded form ▼ |

💊 Save results to a Binder
⁇ Search Tips
☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 1 of 1

Relevance scale ☐▱▰▪■

1 Enhancing performance of protein name recognizers using collocation
Wen-Juan Hou, Hsin-Hsi Chen
July 2003 **Proceedings of the ACL 2003 workshop on Natural language processing in biomedicine - Volume 13**
Publisher: Association for Computational Linguistics
Full text available: 📄 pdf(159.32 KB)   Additional Information: full citation, abstract, references, citings

Named entity recognition is a fundamental task in biological relationship mining. This paper employs protein collocates extracted from a biological corpus to enhance the performance of protein name recognizers. Yapex and KeX are taken as examples. The precision of Yapex is increased from 70.90% to 81.94% at the low expense of recall rate (i.e., only decrease 2.39%) when collocates are incorporated. We also integrate the results proposed by Yapex and KeX, and employs collocates to filter the merg ...

Results 1 - 1 of 1

P⬡RTAL
USPTO

THE ACM DIGITAL LIBRARY

Feedback Report a problem Satisfaction survey

Terms used
query stopwords results compare stop phrases detect

Found **52** of **197,895**

| Sort results by | relevance ▼ |
| Display results | expanded form ▼ |

❤ Save results to a Binder
🔢 Search Tips
☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 52   Result page: **1** 2 3   next

Relevance scale ☐ ▱ ▰ ▰ ■

■

1 Document detection: Inquery system overview ·
John Broglio, James P. Callan, W. Bruce Croft
September 1993 **Proceedings of a workshop on held at Fredericksburg, Virginia: September 19-23, 1993**
Publisher: Association for Computational Linguistics
Full text available: 📄 pdf(1.57 MB)   Additional Information: full citation, abstract, references

> The TIPSTER project in the Information Retrieval Laboratory of the Computer Science Department, University of Massachusetts, Amherst (which includes MCC as a subcontractor), has focused on the following goals:• Improving the effectiveness of information retrieval techniques for large, full-text databases,• Improving the effectiveness of routing techniques appropriate for long-term information needs, and• Demonstrating the effectiveness of these retrieval and routing techniques for ...

2 Internet data management (IDM): Learning query languages of Web interfaces   ■
André Bergholz, Boris Chidlovskii
March 2004 **Proceedings of the 2004 ACM symposium on Applied computing SAC '04**
Publisher: ACM Press
Full text available: 📄 pdf(253.16 KB)   Additional Information: full citation, abstract, references

> This paper studies the problem of automatic acquisition of the query languages supported by a Web information resource. We describe a system that automatically probes the search interface of a resource with a set of test queries and analyses the returned pages to recognize supported query operators. The automatic acquisition assumes the availability of the number of matches the resource returns for a submitted query. The match numbers are used to train a learning system and to generate classific ...

Keywords: hidden Web, learning, query operators, search interface

3 Paper session IR-11 (information retrieval): novelty detection: Novelty detection   ■
based on sentence level patterns
Xiaoyan Li, W. Bruce Croft
October 2005 **Proceedings of the 14th ACM international conference on Information and knowledge management CIKM '05**
Publisher: ACM Press
Full text available: 📄 pdf(94.91 KB)   Additional Information: full citation, abstract, references, index terms

The detection of new information in a document stream is an important component of many potential applications. In this paper, a new novelty detection approach based on the identification of sentence level patterns is proposed. Given a user's information need, some patterns in sentences such as combinations of query words, named entities and phrases, may contain more important and relevant information than single words. Therefore, the proposed novelty detection approach focuses on the identifica ...

Keywords: information patterns, named entities, novelty detection

4   Evaluating the technologies: the Text REtrieval Conferences (TREC): The Text REtrieval Conferences (TRECs)
Donna Harman
May 1996 **Proceedings of a workshop on held at Vienna, Virginia: May 6-8, 1996**
Publisher: Association for Computational Linguistics
Full text available: pdf(3.20 MB)   Additional Information: full citation, abstract, references

There have been four Text REtrieval Conferences (TRECs); TREC-1 in November 1992, TREC-2 in August 1993, TREC-3 in November 1994 and TREC-4 in November 1995. The number of participating systems has grown from 25 in TREC-1 to 36 in TREC-4, including most of the major text retrieval software companies and most of the universities doing research in text retrieval (see table for some of the participants). The diversity of the participating groups has ensured that TREC represents many different appro ...

5   An evaluation of query processing strategies using the TIPSTER collection
James P. Callan, W. Bruce Croft
July 1993 **Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '93**
Publisher: ACM Press
Full text available: pdf(942.61 KB)   Additional Information: full citation, abstract, references, citings, index terms

The TIPSTER collection is unusual because of both its size and detail. In particular, it describes a set of information needs, as opposed to traditional queries. These detailed representations of information need are an opportunity for research on different methods of formulating queries. This paper describes several methods of constructing queries for the INQUERY information retrieval system, and then evaluates those methods on the TIPSTER document collection. Both AdHoc and Routing query ...'

6   Inverted files for text search engines
Justin Zobel, Alistair Moffat
July 2006 **ACM Computing Surveys (CSUR)**, Volume 38 Issue 2
Publisher: ACM Press
Full text available: pdf(944.29 KB)   Additional Information: full citation, abstract, references, index terms

The technology underlying text search engines has advanced dramatically in the past decade. The development of a family of new index representations has led to a wide range of innovations in index storage, index construction, and query evaluation. While some of these developments have been consolidated in textbooks, many specific techniques are not widely known or the textbook descriptions are out of date. In this tutorial, we introduce the key techniques in the area, describing both a core impl ...

Keywords: Inverted file indexing, Web search engine, document database, information retrieval, text retrieval

7

Evaluating the technologies: The text retrieval conferences (TRECS)

Ellen M. Voorhees, Donna Harman
October 1998 **Proceedings of a workshop on held at Baltimore, Maryland: October 13-15, 1998**
Publisher: Association for Computational Linguistics
Full text available: 📄 pdf(2.76 MB)   Additional Information: full citation, abstract, references, citings

> Phase III of the TIPSTER project included three workshops for evaluating document detection (information retrieval) projects: the fifth, sixth and seventh Text REtrieval Conferences (TRECs). This work was co-sponsored by the National Institute of Standards and Technology (NIST), and included evaluation not only of the TIPSTER contractors, but also of many information retrieval groups outside of the TIPSTER project. The conferences were run as workshops that provided a forum for participating gro ...

8 Mining the web: Finding advertising keywords on web pages

Wen-tau Yih, Joshua Goodman, Vitor R. Carvalho
May 2006 **Proceedings of the 15th international conference on World Wide Web WWW '06**
Publisher: ACM Press
Full text available: 📄 pdf(194.37 KB)   Additional Information: full citation, abstract, references, index terms

> A large and growing number of web pages display contextual advertising based on keywords automatically extracted from the text of the page, and this is a substantial source of revenue supporting the web today. Despite the importance of this area, little formal, published research exists. We describe a system that learns how to extract keywords from web pages for advertisement targeting. The system uses a number of features, such as term frequency of each potential keyword, inverse document frequ ...

> **Keywords**: advertising, information extraction, keyword extraction

9 Preparing heterogeneous XML for full-text search

Miro Lehtonen
October 2006 **ACM Transactions on Information Systems (TOIS)**, Volume 24 Issue 4
Publisher: ACM Press
Full text available: 📄 pdf(228.25 KB)   Additional Information: full citation, abstract, references, index terms

> XML retrieval is facing new challenges when applied to heterogeneous XML documents, where next to nothing about the document structure can be taken for granted. We have developed solutions where some of the heterogeneity issues are addressed. Our fragment selection algorithm selectively divides a heterogeneous document collection into equi-sized fragments with full-text content. If the content is considered too data-oriented, it is not accepted. The algorithm needs no information about element n ...

> **Keywords**: XML retrieval, heterogeneous documents, indexing

10 Building efficient and effective metasearch engines

Weiyi Meng, Clement Yu, King-Lup Liu
March 2002 **ACM Computing Surveys (CSUR)**, Volume 34 Issue 1
Publisher: ACM Press
Full text available: 📄 pdf(416.07 KB)   Additional Information: full citation, abstract, references, citings, index terms

> Frequently a user's information needs are stored in the databases of multiple search engines. It is inconvenient and inefficient for an ordinary user to invoke multiple search engines and identify useful documents from the returned results. To support unified

access to multiple search engines, a metasearch engine can be constructed. When a metasearch engine receives a query from a user, it invokes the underlying search engines to retrieve useful information for the user. Metasearch engines have ...

Keywords: Collection fusion, distributed collection, distributed information retrieval, information resource discovery, metasearch

11 Detection and evidence: Improving novelty detection for general topics using
sentence level information patterns
Xiaoyan Li, W. Bruce Croft
November 2006 **Proceedings of the 15th ACM international conference on Information and knowledge management CIKM '06**
Publisher: ACM Press
Full text available: pdf(338.83 KB)   Additional Information: full citation, abstract, references, index terms

The detection of new information in a document stream is an important component of many potential applications. In this work, a new novelty detection approach based on the identification of sentence level *information patterns* is proposed. First, the information-pattern concept for novelty detection is presented with the emphasis on new information patterns for *general topics* (queries) that cannot be simply turned into specific questions whose answers are specific named entities (NE ...

Keywords: information patterns, named entities, novelty detection

12 Information retrieval: Enhancing detection through linguistic indexing and topic expansion
Tomek Strzalkowski, Gees C. Stein, G. Bowden Wise
October 1998 **Proceedings of a workshop on held at Baltimore, Maryland: October 13-15, 1998**
Publisher: Association for Computational Linguistics
Full text available: pdf(973.36 KB)   Additional Information: full citation, abstract, references

Natural language processing techniques may hold a tremendous potential for overcoming the inadequacies of purely quantitative methods of text information retrieval. Under the Tipster contracts in phases I through III, GE group has set out to explore this potential through development and evaluation of new text processing techniques. This work resulted in some significant advances and in a better understanding on how NLP may benefit IR. Tipster research has laid a critical groundwork for future w ...

13 Making MIRACLEs: Interactive translingual search for Cebuano and Hindi
Daqing He, Douglas W. Oard, Jianqiang Wang, Jun Luo, Dina Demner-Fushman, Kareem Darwish, Philip Resnik, Sanjeev Khudanpur, Michael Nossal, Michael Subotin, Anton Leuski
September 2003 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 2 Issue 3
Publisher: ACM Press
Full text available: pdf(209.29 KB)   Additional Information: full citation, abstract, references, index terms

Searching is inherently a user-centered process; people pose the questions for which machines seek answers, and ultimately people judge the degree to which retrieved documents meet their needs. Rapid development of interactive systems that use queries expressed in one language to search documents written in another poses five key challenges: (1) interaction design, (2) query formulation, (3) cross-language search, (4) construction of translated summaries, and (5) machine translation. This articl ...

Keywords: Cross-language information retrieval, Interactive information retrieval, .

Machine translation

14 Embedding web-based statistical translation models in cross-language information retrieval

Wessel Kraaij, Jian-Yun Nie, Michel Simard
September 2003 **Computational Linguistics,** Volume 29 Issue 3
Publisher: MIT Press

Full text available: pdf(381.29 KB) Additional Information: full citation, abstract, references, citings, index terms

Although more and more language pairs are covered by machine translation (MT) services, there are still many pairs that lack translation resources. Cross-language information retrieval (CLIR) is an application that needs translation functionality of a relatively low level of sophistication, since current models for information retrieval (IR) are still based on a bag of words. The Web provides a vast resource for the automatic construction of parallel corpora that can be used to train statistical ...

15 Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies

Soumen Chakrabarti, Byron Dom, Rakesh Agrawal, Prabhakar Raghavan
August 1998 **The VLDB Journal — The International Journal on Very Large Data Bases,** Volume 7 Issue 3
Publisher: Springer-Verlag New York, Inc.

Full text available: pdf(281.37 KB) Additional Information: full citation, abstract, citings, index terms

We explore how to organize large text databases hierarchically by topic to aid better searching, browsing and filtering. Many corpora, such as internet directories, digital libraries, and patent databases are manually organized into topic hierarchies, also called *taxonomies*. Similar to indices for relational data, taxonomies make search and access more efficient. However, the exponential growth in the volume of on-line textual information makes it nearly impossible to maintain such taxono ...

16 Building effective queries in natural language information retrieval

Tomek Strzalkowski, Fang Lin, Jose Perez-Carballo, Jin Wang
March 1997 **Proceedings of the fifth conference on Applied natural language processing**
Publisher: Morgan Kaufmann Publishers Inc.

Full text available: pdf(771.03 KB)
Publisher Site   Additional Information: full citation, abstract, references, citings

In this paper we report on our natural language information retrieval (NLIR) project as related to the recently concluded 5th Text Retrieval Conference (TREC-5). The main thrust of this project is to use natural language processing techniques to enhance the effectiveness of full-text document retrieval. One of our goals was to demonstrate that robust if relatively shallow NLP can help to derive a better representation of text documents for statistical search. Recently, we have turned our attenti ...

17 Query clustering using user logs

January 2002 **ACM Transactions on Information Systems (TOIS),** Volume 20 Issue 1
Publisher: ACM Press

Full text available: pdf(1.31 MB)   Additional Information: full citation, abstract, references, citings, index terms, review

Query clustering is a process used to discover frequently asked questions or most popular topics on a search engine. This process is crucial for search engines based on question-

answering. Because of the short lengths of queries, approaches based on keywords are not suitable for query clustering. This paper describes a new query clustering method that makes use of user logs which allow us to identify the documents the users have selected for a query. The similarity between two queries may be ded ...

Keywords: Query clustering, search engine, user log, web data mining

18   Writing the web: Mining topic-specific concepts and definitions on the web

Bing Liu, Chee Wee Chin, Hwee Tou Ng
May 2003   **Proceedings of the 12th international conference on World Wide Web WWW '03** .
Publisher: ACM Press

Full text available: pdf(245.66 KB)   Additional Information: full citation, abstract, references, citings, index terms

Traditionally, when one wants to learn about a particular topic, one reads a book or a survey paper. With the rapid expansion of the Web, learning in-depth knowledge about a topic from the Web is becoming increasingly important and popular. This is also due to the Web's convenience and its richness of information. In many cases, learning from the Web may even be essential because in our fast changing world, emerging topics appear constantly and rapidly. There is often not enough time for someone ...

Keywords: definition mining, domain concept mining, information integration, knowledge compilation, web content mining

19   Clustering user queries of a search engine

Ji-Rong Wen, Jian-Yun Nie, Hong-Jiang Zhang
April 2001   **Proceedings of the 10th international conference on World Wide Web WWW '01**
Publisher: ACM Press
Full text available: pdf(219.35 KB)   Additional Information: full citation, references, citings, index terms

Keywords: query clustering, search engine, user log, web data mining

20   A highly scalable and effective method for metasearch

Weiyi Meng, Zonghuan Wu, Clement Yu, Zhuogang Li
July 2001   **ACM Transactions on Information Systems (TOIS)**, Volume 19 Issue 3
Publisher: ACM Press

Full text available: pdf(653.63 KB)   Additional Information: full citation, abstract, references, citings, index terms

A metasearch engine is a system that supports unified access to multiple local search engines. Database selection is one of the main challenges in building a large-scale metasearch engine. The problem is to efficiently and accurately determine a small number of potentially useful local search engines to invoke for each user query. In order to enable accurate selection, metadata that reflect the contents of each search engine need to be collected and used. This article proposes a highly scalable ...

Keywords: Database selection, distributed text retrieval, metasearch engine, resource discovery
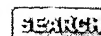
Results 1 - 20 of 52   Result page: **1** 2 3   next

P⊘RTAL
USPTO

THE ACM DIGITAL LIBRARY

¶ Feedback Report a problem Satisfaction survey

Terms used
query stopwords results compare stop phrases detect

Found **3** of **197,895**

Sort results by   [relevance ▾]   ❧ Save results to a Binder   Try an Advanced Search
                                        [?] Search Tips   Try this search in The ACM Guide
Display results   [expanded form ▾]   ☐ Open results in a new window

Results 1 - 3 of 3

Relevance scale ☐ ◱ ◰ ◼ ■

1 Document detection: Inquery system overview   ■
John Broglio, James P. Callan, W. Bruce Croft
September 1993 **Proceedings of a workshop on held at Fredericksburg, Virginia: September 19-23, 1993**
Publisher: Association for Computational Linguistics
Full text available: 🗎 pdf(1.57 MB)   Additional Information: full citation, abstract, references

> The TIPSTER project in the Information Retrieval Laboratory of the Computer Science Department, University of Massachusetts, Amherst (which includes MCC as a subcontractor), has focused on the following goals:• Improving the effectiveness of information retrieval techniques for large, full-text databases,• Improving the effectiveness of routing techniques appropriate for long-term information needs, and• Demonstrating the effectiveness of these retrieval and routing techniques for ...

2 Embedding web-based statistical translation models in cross-language information   ■
retrieval
Wessel Kraaij, Jian-Yun Nie, Michel Simard
September 2003 **Computational Linguistics**, Volume 29 Issue 3
Publisher: MIT Press
Full text available: 🗎 pdf(381.29 KB)   Additional Information: full citation, abstract, references, citings, index terms

> Although more and more language pairs are covered by machine translation (MT) services, there are still many pairs that lack translation resources. Cross-language information retrieval (CLIR) is an application that needs translation functionality of a relatively low level of sophistication, since current models for information retrieval (IR) are still based on a bag of words. The Web provides a vast resource for the automatic construction of parallel corpora that can be used to train statistical ...

3 An evaluation of query processing strategies using the TIPSTER collection   ■
James P. Callan, W. Bruce Croft
July 1993 **Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '93**
Publisher: ACM Press
Full text available: 🗎 pdf(942.61 KB)   Additional Information: full citation, abstract, references, citings, index terms

> The TIPSTER collection is unusual because of both its size and detail. In particular, it

describes a set of information needs, as opposed to traditional queries. These detailed representations of information need are an opportunity for research on different methods of formulating queries. This paper describes several methods of constructing queries for the INQUERY information retrieval system, and then evaluates those methods on the TIPSTER document collection. Both AdHoc and Routing query ...

Results 1 - 3 of 3